A phonetically-guided diagnosis of auditory deficiency based on synthetic speech stimuli

Anne Bonneau, Parham Mokhtari

A PHONETICALLY-GUIDED DIAGNOSIS OF AUDITORY DEFICIENCY BASED ON SYNTHETIC SPEECH STIMULI

Anne Bonneau, Parham Mokhtari
LORIA/CNRS and INRIA, BP 239, 54506 Vandœuvre-lès-Nancy, FRANCE.
E-mail:bonneau@loria.fr, parham@etl.go.jp

ABSTRACT

We propose a phonetically-guided diagnosis of auditory deficiency, which hinges on a carefully constructed corpus of synthetic sounds. Our aim is to complement the diagnosis of sensorineural hearing deficiencies, in order to improve the correction afforded by auditory prosthesis. We first design a vowel corpus, based upon pairs of two-formant, steady-state synthetic French vowels, where the vowels of each pair are chosen to differ only in the frequency of one of their two formants. To test our method, we simulate a frequency-selective loss of audibility, by specifying a piece-wise linear audibility curve with a minimum of -40 dB at a given centre-frequency (1.3, 1.6, and 1.9 kHz). Results of perceptual experiments with normal hearing people tend to show that our synthetic data set is amenable to the diagnosis of the frequency region where the simulated hearing problem is most acute.
Keywords : audioprosthesis, synthetic vowels, diagnosis.

1. INTRODUCTION

Speech understanding difficulties of people with hearing impairment are known to arise from a number of intertwined, auditory-perceptual deficiencies. Amongst the most important perceptual consequences of sensorineural hearing impairment in particular (which is associated with deficiencies of cochlear origin) are the frequency-dependent rise in the threshold of audibility, a reduced degree of frequency selectivity, and a reduced degree of temporal resolution [1]. Progress in our understanding of the mechanisms which underlie those perceptual consequences, together with the development of speech signal processing techniques, will allow the implementation of more effective signal transformations in audioprosthesis or cochlear implants [2].

Indeed, our speech research team, in collaboration with two medical teams specialized in audiology, is working on speech signal transformations aimed at making more "precise" corrections to auditory deficiencies. Such corrections are specifically adapted to each subject's deficiencies, while paying particular attention to phonetic oppositions between speech sounds. Amongst the possible transformations devoted to the hearing deficiencies mentioned above, we are particularly interested in the enhancement of formants or other important acoustic cues, as well as a selective slowing of the speech rate.

Our aim in this paper is to complement the diagnosis of sensorineural hearing deficiencies, in order to improve the correction afforded by auditory prosthesis. To date, there are at least three main types of diagnosis: tonal audiometry, speech-based audiometry, and psychoacoustic tests. The first of these is limited to evaluating the frequency-dependent loss of sensitivity to the amplitude of tonal signals, and therefore does not directly provide an indication of other types of deficiency. On the other hand, while psychoacoustic tests can provide quantitative measures of deficiencies such as loss of frequency-selectivity, they are often lengthy and tiring for the patient. Of particular interest therefore is speech-based audiometry, which involves the more natural task of speech recognition, and thus bears more directly on the main purpose of auditory prostheses: that of improving speech intelligibility.

In this vein, we here propose a new, phonetically-guided diagnosis of auditory deficiency. Our diagnosis hinges on a reasonably small battery of speech-based auditory tests, based on a carefully constructed corpus of synthetic vowel sounds. Synthetic stimuli are preferred, as they afford control of auditorily-relevant acoustic parameters such as the frequencies and amplitudes of formants. Our synthetic corpus consists of pairs of two- formant, steady-state French vowels, where the two vowels of each pair are carefully chosen to differ only in the frequency of one of their two formants. The aim of using such minimally-contrastive pairs is to better identify those spectral regions in which the patient's auditory deficiencies are the most acute.

In section 2 we describe our experimental methods and sets of vowel stimuli, and in section 3 we present computer simulations designed to test our proposed diagnosis. The results are presented and discussed in section 4.

2. STIMULI AND METHODS

2.1. The vowel corpus

Our corpus of synthetic stimuli for diagnosis of hearing-impairment is made up of two-formant, French

vowels. The formant frequencies of these vowels were chosen carefully in order to: (i) ensure that they cover a wide range of the F1-F2 vocalic space, and (ii) secure minimally- contrastive pairs (and one quadruplet) of vowels which differ only in the frequency of one of their two formants. Owing to the clear formant structure of our vocalic stimuli, the patterns of confusions amongst the vowels are then used (in section 4) to identify those spectral regions in which the patient's auditory deficiencies are the most acute.

All the stimuli were generated using the parallel branch of the Klatt synthesizer. The duration of each stimulus was fixed at 200 ms. The formant frequencies for the male set are listed in Table 1. The fundamental frequency was set to 120 Hz for the male, and 220 Hz for the female voice; however, in order to improve the naturalness of the synthetic stimuli, a linearly downward- sloping F0 profile was generated, with a total drop of 5% from the beginning to the end of each stimulus. Furthermore, the profile of energy was made to fade-in at the start and fade-out at the end of each stimulus, thereby avoiding any clicks or sharp discontinuities in overall level.

Finally, the stimuli were replicated at several different levels of additive noise, which is known to be harmful to speech intelligibility for the hearing-impaired in particular.

2.2. Presentation of stimuli

The synthetic vowel corpus described above was organised into five sets of presentation conditions. In the first four of these, the stimuli with male characteristics were presented under the following conditions:

i. binaurally, without addition of noise (first set)
ii. binaurally, with addition of noise (second set)
iii. dichotically, without addition of noise (third set)
iv. dichotically, with addition of noise (fourth set),

where "dichotic" implies presentation of the F1 of a given stimulus to one ear, while simultaneously presenting its F2 to the other ear; and the additive noise was a white noise.

The stimuli of the fifth set were those synthesised with female characteristics, and were presented only binaurally, without addition of noise.

Perceptual experiments [3] do suggest that vowel timbre is integrated at a non-peripheral level, since the presentation of one or more formants to one ear and the remaining formants to the other ear was found not to be harmful to the perceptual identification of the vowel. Furthermore, Chaudari *et al.* (1998) [4] simulated sensorineural hearing loss of frequency selectivity and were able to demonstrate that the dichotic presentation of split stimuli improved consonant perception. Indeed, the efficacy of dichotic presentation of our stimuli (F1 spectral information in one ear, that of F2 in the other ear) will next be evaluated in terms of its potential for alleviating those vowel-intelligibility problems which may be caused mainly by reduced degrees of frequency selectivity.

3. PERCEPTUAL TESTS

The experimental methodology adopted to test our proposed method of diagnosis, involves first computer simulations of certain aspects of hearing impairment, then presentation of thus generated stimuli to normal-hearing listeners. We present below the experimental protocol, the results will be presented in the next section.

3.1. Experimental protocol

The listeners used in all the experiments reported in this paper, were ten, native speakers of French, with no known hearing impairment. They listened to the stimuli in a quiet room using Sennheiser HD520 II headphones, with the volume control adjusted to a comfortable level.

They were asked to choose their response from amongst the 6 French vowels /i, E, A, y, OE, O/, where /E/,/A/,/OE/, /O/ represent /e, ɛ/, /a, ɑ/, /ø, œ/ and /o, ɔ/ respectively. This list of vowels intentionally excludes the vowels not present in our alphabet (such as /u/) as well as distinctions between pairs of vowels whose phonological opposition are often neutralized in French. These distinctions were excluded from our proposed method of hearing-impairment diagnosis, on grounds of their well-known difficulty, even for normal-hearing, French-speaking listeners.

The listeners responded orally after two consecutive presentations of each stimulus, separated by a one-second interval. A four-second interval was then used to separate two different stimuli.

3.2. Identification of the unmodified stimuli

We first tested whether the original vowel stimuli (i.e., those not subjected to modifications or to additive noise) could be well identified by the listeners. To this end, we presented the subjects with the following, three sets of vowel stimuli: stimuli with low F0 (first set as described earlier), with high F0 (fifth set), and with low F0 under dichotic conditions (third set). Within each set, the stimuli were presented in a random order, and each stimulus was presented twice.

3.3. Simulation of hearing impairment

We first simulated a frequency-selective loss of audibility, by specifying a piece-wise linear audibility curve with a minimum of -40 dB at a given centre-frequency (1300, 1600, then 1900 Hz), linear segments joining that point to 0 dB at 500 Hz either side of the centre-frequency, and 0 dB elsewhere along the available frequency range. The inverted-triangular audibility curve thus obtained, was then added to every short-time FFT spectrum of the stimulus, and the resulting, modified spectra were used to synthesise the new stimulus by the well-known OverLap-Add (OLA) method.

These simulations were performed on the first set of stimuli (male voice). We thus generated three new sets of

stimuli, one for each center-frequency. We presented the three sets separately to the subjects. As in the previous test, each stimulus was presented twice, and the stimuli within each set were randomized.

4. RESULTS

4.1. Identification of the unmodified stimuli

As shown in Table 2, the identification rates of the original, unmodified stimuli were consistently equal to or above 75%, with the exception of the vowel /ɔ/ (low F0) which was confused with the vowel /oe/ (a confusion which occurs frequently, even for a natural, isolated /ɔ/). Indeed, an analysis of the individual vowel confusions revealed that they all occur between vowels very close together in the vocalic space (e.g., /i,e/).

It is interesting to note, however, that most of the vowels in the female set were quite perfectly intelligible. We also observe an overall increase in the intelligibility of the vowels in the male set, when they are presented dichotically; this is particularly evident for the front unrounded vowels, and the vowel /ɔ/. Further experiments would be necessary to verify this result. In future experiments, we plan to improve the quality of our vowel stimuli (especially those in the male set) by finding the perceptually "preferred" F2 value for each timbre, thereby raising our benchmark identification rates to above 85%.

4.2. Identification of stimuli with simulated hearing-impairment

Previous experiments [5] with synthetic, two-formant vowels have shown that a sufficient decrease in the amplitude of the second formant of a front vowel can lead to the perception of the corresponding back vowel (i.e., that which shares a similar F1). On the other hand, the presence of F1 alone is generally sufficient for the identification of back vowels. Hence, a decrease in the intelligibility of front vowels after simulation of the inverted-triangular audibility curve as described in section 3.3, should indicate a problem in the region of the F2 of the vowels thus badly perceived, especially if another (non-back) vowel with a similar F1 is not (or is less) affected. With this type of reasoning, our well-controlled, synthetic data set is expected to be amenable to the diagnosis of the frequency region where the simulated hearing problem is most acute.

Indeed, as described below, our results tend to confirm this type of diagnosis.

As shown in Figure 1, the vowels that were most affected by our simulation of audibility-loss centered at 1300 Hz were those central and front vowels with F2 equal to or slightly higher than 1300 Hz. Similarly, our simulation of audibility-loss centered at 1600 Hz detrimentally affected the intelligibility of mainly /ø/ (F2 at 1600 Hz), but also that of /y/ (F2 at 1800 Hz). Simulation of audibility-loss centered at 1800 Hz strongly affected the intelligibility of the vowel /y/ (F2 at 1800 Hz). Our results therefore clearly show that the frequencies of the second formants of the most affected vowels do provide an indication of the frequency range in which maximal loss of audibility was simulated.

At the same time, however, it is worth mentioning that certain vowels with F2 quite close to the centre-frequency of simulated loss were still well-identified perceptually. We had earlier mentioned this to be the case for the back vowels, where F1 is known to be of primary perceptual importance. However, we now see a similar phenomenon for the front vowel /ɛ/ which, unlike the front and rounded vowel /y/ which shares the same F2 of 1800 Hz, was not strongly affected by our simulations of audibility loss at 1600 Hz, nor at 1900 Hz. An explanation of this phenomenon requires further investigations, and is beyond the scope of the present article.

5. CONCLUSION

We have presented a new, phonetically-guided diagnosis of auditory deficiency, based upon pairs of two-formant, steady-state synthetic French vowels, where the vowels of each pair are carefully chosen to differ only in the frequency of one of their two formants. Our (very simple) simulations of frequency-selective losses of audibility, tend to confirm that our well-controlled, synthetic data set is amenable to the diagnosis of the frequency region where the simulated hearing problem is most acute.

Our future investigations will concern the simulations of reduced degrees of frequency selectivity, the efficacy of the dichotic presentation, as well as the effect of different levels of additive noise on the perception of our modified and unmodified stimuli. The corpus will then be tested on hearing impaired people at the Central Hospital of Nancy, with the help of specialists in speech-based audiometry. In addition, we intend to complement our corpus with synthetic consonants in vocalic environments.

6. REFERENCES

[1]Moore, B.C.J. (1995), Perceptual consequences of cochlear damage. Oxford University Press.

[2]Loizou, P.C. (1998), Mimicking the human ear. *IEEE Signal Processing Magazine*, pp 101-130.

[3]Carlson, R. Fant, G. and Granstrom (1975), Two formant models, pitch and vowel perception. In G. Fant and M.A. Tatham, editors, *Auditory analysis and perception of speech*, pp 55-82. Academic Press, New-York.

[4]Chaudari, D.S. and Pandey, P.C. (1998), Dichotic presentation of speech signal with critical band filtering for improving speech perception. *Proceedings of ICASSP, Berlin*.

[5]Ainsworth, W. A. and Millar, J.B. (1972), The effect of relative formant amplitude on the perceived identity of synthetic vowels. *Language and Speech*, 15:328-341.

F1-F2	900	1050	1300	1600	1800	2100	3000
300					y		i
380				ø		e	
500	o	ɔ	œ		ɛ		
650		ɑ	a				

Table 1: Formant frequencies for the male set of synthetic vowels. It is to be noted that for each F1, there exist at least two vowels (four at F1 = 500 Hz) differing only in F2; conversely, there are three pairs of vowels which differ only in F1.

	i	e	ɛ	a	y	ø	œ	ø	ɔ	ɑ
M	95	75	85	100	90	75	100	100	65	90
F	100	75	75	100	100	100	100	100		
Dic	100	100	90	100	85	100	90	90	90	85

Table 2: Vowel identification rates (%) for the three sets of unmodified stimuli: the male (M), the female set (F) and the male set presented dichotically (Dic).

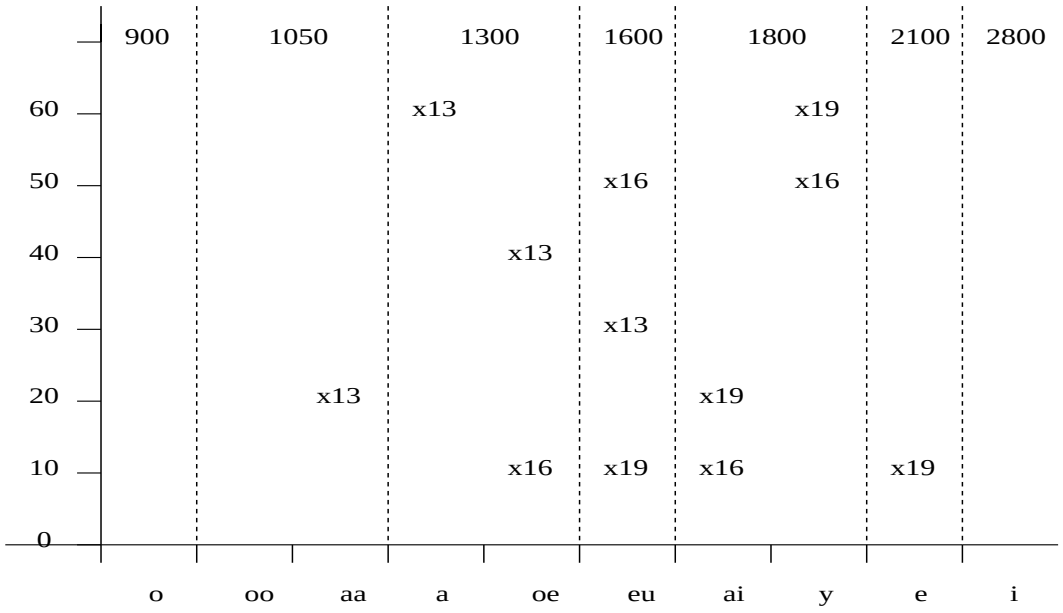

Figure 1: Decrease (%) in vowel identification rates after simulation of loss-of-audibility centered at 1300 (x13) 1600 (x16) and 1900 Hz (x19). The symbols represent the vowels /o,ɔ,ɑ,a,œ,ø,ɛ,y,e,i/. The vertical dotted lines separate the vowels with different F2, whose values are indicated along the top of the figure.